

TarBase-v9.0 extends experimentally supported miRNA–gene interactions to cell-types and virally encoded miRNAs

Giorgos Skoufos^{1,2,*†}, Panos Kakoulidis^{3,4,†}, Spyros Tastsoglou^{1,2,†}, Elissavet Zacharopoulou^{1,2}, Vasiliki Kotsira^{1,2}, Marios Miliotis^{1,2}, Galatea Mavromati¹, Dimitris Grigoriadis¹, Maria Zioga¹, Angeliki Velli¹, Ioanna Koutou¹, Dimitra Karagkouni^{1,2}, Steve Stavropoulos⁵, Filippos S. Kardaras^{1,2}, Anna Lifousi⁶, Eustathia Vavalou⁷, Armen Ovsepian^{1,2}, Anargyros Skoulakis^{1,2}, Sotiris K. Tasoulis⁵, Spiros V. Georgakopoulos⁸, Vassilis P. Plagianakos⁵ and Artemis G. Hatzigeorgiou^{1,2,*}

¹DIANA-Lab, Dept. of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

²Hellenic Pasteur Institute, Athens 11521, Greece

³Dept. of Informatics and Telecommunications, National and Kapodistrian Univ. of Athens, Athens, Greece

⁴Biomedical Research Foundation of the Academy of Athens, 11527 Athens, Greece

⁵Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

⁶Technical University of Denmark – Department of Health Technology, Copenhagen, Denmark

⁷Department of Biology, National and Kapodistrian University of Athens, 15784 Athens, Greece

⁸Department of Mathematics, University of Thessaly, Greece

*To whom correspondence should be addressed. Tel: +302421074758; Fax: +30 24210 74997; Email: arhatzig@uth.gr

Correspondence may also be addressed to Giorgos Skoufos. Tel: +302421074758; Fax: +302421074997; Email: gskoufos@uth.gr

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Present addresses:

Dimitra Karagkouni, Cancer Research Institute | Harvard Medical School, Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA.

Dimitra Karagkouni, Harvard Medical School, Boston, MA, USA.

Abstract

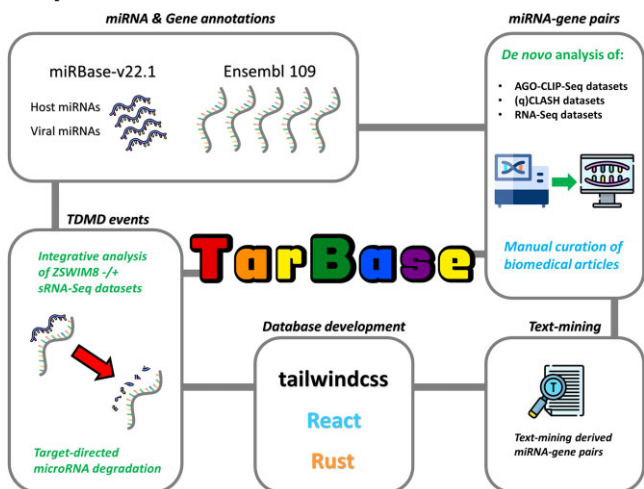
TarBase is a reference database dedicated to produce, curate and deliver high quality experimentally-supported microRNA (miRNA) targets on protein-coding transcripts. In its latest version (v9.0, <https://dianalab.e-ce.uth.gr/tarbasev9>), it pushes the envelope by introducing virally-encoded miRNAs, interactions leading to target-directed miRNA degradation (TDMD) events and the largest collection of miRNA–gene interactions to date in a plethora of experimental settings, tissues and cell-types. It catalogues ~6 million entries, comprising ~2 million unique miRNA–gene pairs, supported by 37 experimental (high- and low-yield) protocols in 172 tissues and cell-types. Interactions are annotated with rich metadata including information on genes/transcripts, miRNAs, samples, experimental contexts and publications, while millions of miRNA-binding locations are also provided at cell-type resolution. A completely re-designed interface with state-of-the-art web technologies, incorporates more features, and allows flexible and ingenious use. The new interface provides the capability to design sophisticated queries with numerous filtering criteria including cell lines, experimental conditions, cell types, experimental methods, species and/or tissues of interest. Additionally, a plethora of fine-tuning capacities have been integrated to the platform, offering the refinement of the returned interactions based on miRNA confidence and expression levels, while boundless local retrieval of the offered interactions and metadata is enabled.

Received: September 15, 2023. Revised: October 18, 2023. Editorial Decision: October 20, 2023. Accepted: November 2, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical abstract



Introduction

miRNAs play important roles in many aspects of molecular biology in both physiological and pathological conditions including cardiovascular and neural development, stem cell differentiation, metabolism, apoptosis, neurodegenerative diseases and tumors (1–5). A large body of evidence from recently published studies suggest that miRNAs also have alternative functions; in some cases, miRNAs have been identified to upregulate protein expression (6), activate transcription by direct interaction with DNA (7), target non-coding RNAs in the cell nucleus (8) and be highly enriched in the extracellular space (9). Their presence in the extracellular space (e.g. blood plasma), inside exosome-like extracellular micro-vesicles, has been extensively studied under the contexts of cell–cell signalling and biomarker potential (10,11). Additionally, evidence suggests that miRNAs may also partake in paracrine signalling, binding Toll-like receptors (12,13).

Since their initial discovery, numerous experimental and computational methods have been developed aiming at the identification, annotation, *de novo* prediction and cataloguing of miRNA targets on coding and non-coding transcripts. AGO-CLIP-Seq protocols (14,15) (e.g. HITS-CLIP, PAR-CLIP) have been widely adopted, offering context-specific miRNA interactomes in a transcriptome-wide manner. Furthermore, CLEAR-CLIP (16), CLASH (17) and qCLASH (18) protocols blaze the trail towards direct observations of miRNA-RNA pairs by utilizing a ligation step that links miRNAs with their RNA targets resulting in chimeric RNA fragments prior to sequencing. Additional methods with focus in the identification of miRNA–gene pairs have been employed; such methods include miRNA transfection/knockout cells followed by Microarray, RNA-Seq, RPF-Seq and RIP-Seq experiments, RT-qPCR, Western blot and Luciferase Reporter Assays and others.

Except for the miRNA-induced degradation and translational suppression of their targets, another major emerging process termed target-directed miRNA degradation (TDMD), promotes the destruction of miRNAs instead (19,20). Mechanistically, TDMD events start by the extensive pairing of the miRNA 3' region to the target RNA molecule; this results in the destabilization of the miRNA which in turn is recognized

by the ZSWIM8 ubiquitin ligase. ZSWIM8 adds polyubiquitin chains to AGO, leading to its proteolysis and subsequently exposing the miRNA to degradation.

The systematic curation and indexing of the rapidly expanding volume of miRNA–gene interactions in different experimental settings and species is vital to basic and applied RNA research and is addressed by TarBase and other online repositories. miRTarBase (21) is a database devoted to miRNA–gene interactions. The database comprises >2 million miRNA–target pairs from low- and high-throughput methodologies through manual curation of the available literature and the analysis of CLIP-Seq datasets. ENCORI/StarBase (22) is an online platform focused on RNA-RNA and protein-RNA interactions. The database contains more than 1 million miRNA-mRNA interactions and additional interactions between other RNA species and proteins. MiRecords (23) is another database with similar content, albeit with infrequent updates (last updated in 2009) and significantly fewer interactions (i.e. ~3000 entries).

We present TarBase-v9.0 (Figure 1), an extensive online repository of ~6 million entries, 3-fold more than miRTarBase and 6-fold more than TarBase-v8.0, making it the largest collection of experimentally supported miRNA–gene interactions to date. It is the first TarBase version to employ microCLIP (24), an *avant-garde* CLIP-Seq analysis framework that combines deep learning classifiers under a super learning scheme for CLIP-Seq-guided detection of miRNA interactions. TarBase-v9.0 also incorporates interactions from direct miRNA–target chimeras produced by the *de novo* analysis of CLASH and qCLASH experiments, as well as miRNA–gene pairs derived from the analysis of miRNA-specific transfection/knockdown RNA-Seq experiments with a standardized data processing pipeline. Importantly, TarBase-v9.0 introduces (i) ~34 000 interacting pairs between host mRNAs and virally-encoded miRNAs, (ii) interactions leading to target-directed miRNA degradation (TDMD) events and (iii) millions of exact miRNA-binding locations with cell-type resolution.

TarBase interface has been entirely revamped, introducing an array of additional features that enhance its versatility and usability. The newly introduced interface empowers users to design sophisticated queries, offering a multitude of filtering

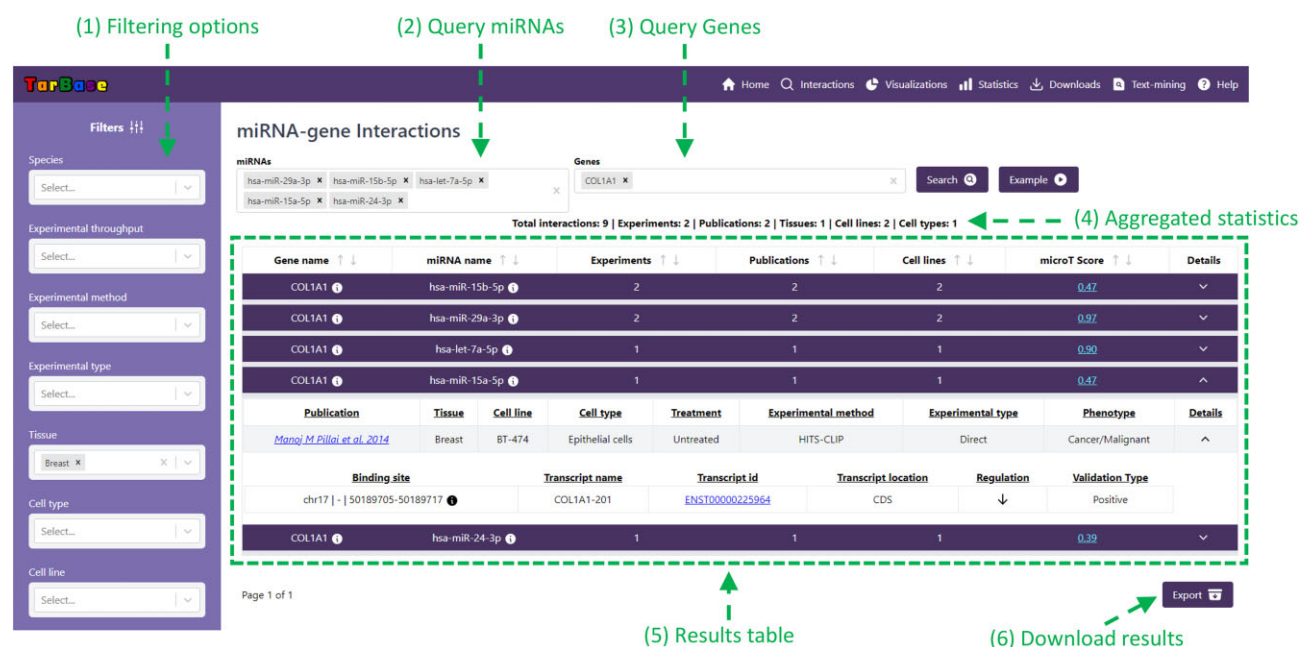


Figure 1. TarBase-v9.0 main user interface. (1) TarBase offers numerous filtering options (i.e. cell lines, experimental conditions, cell types etc.). Users may conduct queries with one or more miRNA(s) (2) and/or one or more genes (3). (4) Aggregated statistics of the returned interactions, (5) Results table and (6) an option to download the returned interactions.

options such as cell lines, experimental conditions, cell types, experimental techniques, species, and tissues of interest. Furthermore, an abundance of fine-tuning capacities has been seamlessly integrated into the platform, enabling users to refine their search results based on miRNA confidence, expression levels, as well as microCLIP and DIANA-microT 2023 (25) scores. Additionally, the platform now facilitates unrestricted local retrieval of both the interactions and all associated metadata.

Methods and results

Database statistics and content

De novo analysis of Next-Generation Sequencing (NGS) protocols including AGO-PAR-CLIP, CLASH and RNA-Seq samples and the manual curation of biomedical literature yielded a total of ~6 million entries (Figure 2). These interactions comprise ~2 million unique miRNA–gene pairs spanning over 103 tissues, 57 cell-types and >300 cell-lines. In total, TarBase features interactions derived from 22 low- and 15 high-throughput experimental techniques. Furthermore, TarBase-v9.0 incorporates >3300 miRNAs and >40 000 genes spreading over 24 species (e.g. *Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster* etc.). Moreover, for every annotated miRNA–gene pair, the database integrates additional information including article metadata, details on genes and miRNAs (e.g. Ensembl (26) and miRBase (27) names and stable identifiers, the miRNA sequence, the targeted region genomic coordinates etc.), sample treatments, phenotypes, microT-CDS-2023 (25) predicted scores and more.

TarBase-v9.0 provides a ~6-fold increase in total and a ~3-fold increase in unique miRNA–gene interactions compared to TarBase-v8.0 (28). Additionally, it features >45 000 candidate Target-Directed miRNA Degradation events which to the best of our knowledge is the first ever biomedical database to host such information in a well curated collection. Finally, for the

first time, boundless local retrieval of the entirety of TarBase interactions and metadata is offered to users.

Database functionality and interface

TarBase-v9.0 interface has been redesigned from the ground up, bringing in numerous new features that improve its flexibility and user-friendliness. The development of the latest TarBase was conducted using state-of-the-art web technologies, enabling users to conduct queries with lightning-fast response times. Interactions can be retrieved by performing queries with miRNA or gene names/identifiers or both simultaneously. A plethora of filtering options to refine the returned miRNA–target pairs can be applied for species, experimental method, experimental type (i.e. direct or indirect protocols), cell line, tissue, cell type and gene region (i.e. 3'UTR or CDS). Additional fine-tuning can be employed with regard to miRNA expression, microT-CDS-2023 (25) prediction scores, miRNA confidence (miRBase-v22.1), binding site information and year of publication.

TarBase-v9.0 is also seamlessly interconnected with other biomedical databases. Since its sixth version, TarBase has been integrated in Ensembl (26). Interactions with exact binding sites can be viewed in the Ensembl genome browser; gene and miRNA information can be accessed through Ensembl and miRBase (27) respectively. TarBase is also included as an expert non-coding RNA resource in the RNACentral knowledgebase. Finally, information on publications can be directly retrieved through the interconnection with PubMed while miRNA–target interaction scores and predicted binding locations can also be reached by DIANA-microT-CDS 2023.

Furthermore, a separate page comprising miRNA–gene interactions (at miRNA precursor level) which have been automatically extracted via an auxiliary text mining pipeline with full-text capacity is incorporated to the interface. Basic statistics of the database (e.g. top experimental methods) are

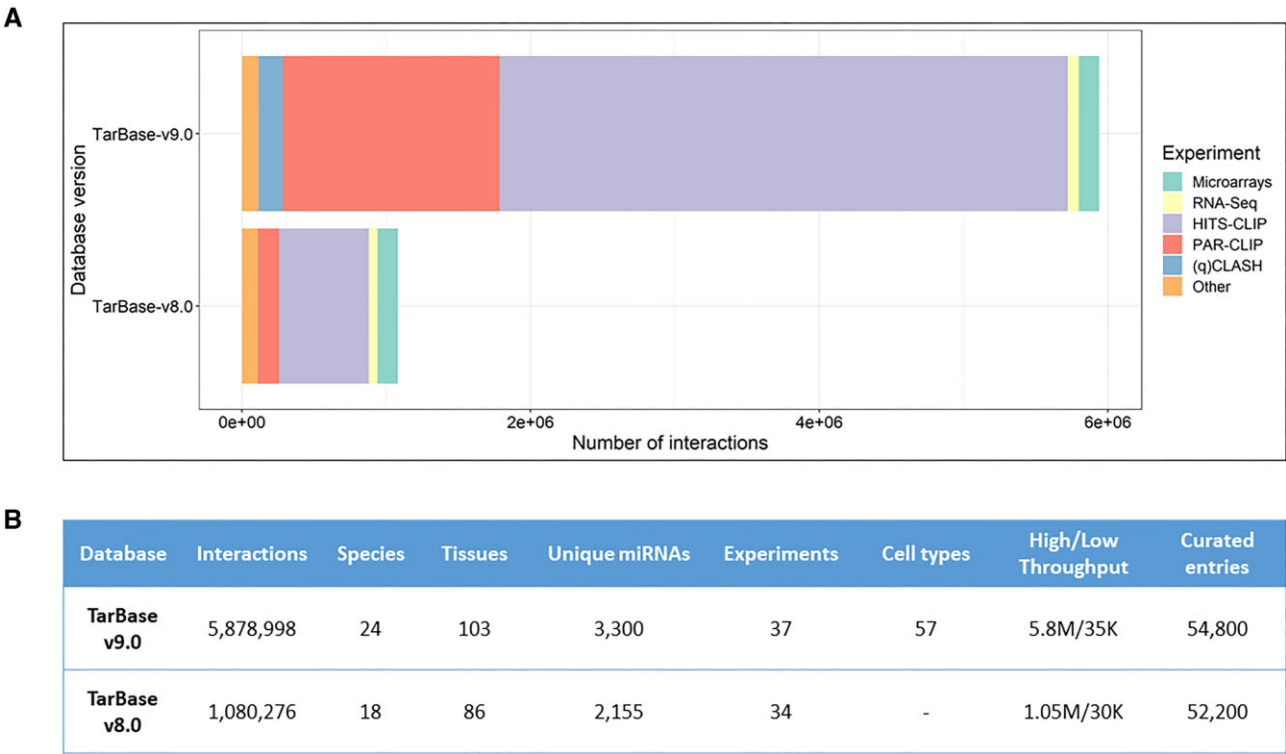


Figure 2. Comparison of TarBase-v9.0 versus TarBase-v8.0 in terms of **(A)** total interactions per experimental method and **(B)** other basic statistics of the databases.

presented through bar-plots in a dedicated page. A visualization page to assess the combinatorial effect of multiple miRNAs in one or multiple genes is also provided; using this feature, users may conduct queries with two or more miRNAs and a visual representation of the interactions between the selected miRNAs and their respected targets will be rendered in the form of a network graph. For a gene to be included in the network graph, it must interact with at least two of the selected miRNAs. Finally, a help page with rich material (screenshots and text-based walkthroughs) is available to guide users and ensure a smooth experience for first-time visitors.

Analysis of high-throughput datasets

AGO-CLIP-Seq. Raw AGO-CLIP-Seq libraries were quality checked and pre-processed using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and cutadapt (29). Subsequently, CLIP-Seq samples were aligned against the human and mouse reference genomes (i.e. GRCh38 and mm10) with GMAP/GSNAP (30). microCLIP (24) algorithm was employed to identify binding events between mRNA peaks and expressed miRNAs. Top expressed miRNAs were retrieved either from (i) the relevant publications, (ii) the analysis of small RNA-Seq (sRNA-Seq) libraries applied to the same cell types/tissues or (iii) databases with miRNA expression profiles across tissues and cell-types (31,32). In datasets with more than one biological replicates, a miRNA binding event had to be present in at least two replicates. Analysis of sRNA-Seq datasets to derive miRNA expression levels was conducted using DIANA-mAP (33).

CLASH. Raw Quick CLASH (qCLASH) and CLASH datasets were retrieved and analyzed using Hyb workflow (34) and RNAup from ViennaRNA package (35) for RNA-RNA

interaction evaluation. For samples infected with viruses, tailored Hyb databases were built integrating host transcripts, host miRNAs and virally encoded miRNAs. Within each study and condition, a miRNA binding event was annotated as replicated if it exhibited the same transcript coordinates in more than one replicate sample. (q)CLASH interactions were thus annotated regarding the total number of supporting chimeric reads across replicates, the estimated Minimum Free Energy (MFE, kcal/mol) of the binding region and the number of replicates supporting each event. Interactions with a negative MFE, supported by >1 total chimeric reads, were retained.

RNA-Seq. Raw RNA-Seq libraries were quality checked and pre-processed using FastQC, minion (36) and cutadapt. Pre-processed reads were aligned to their respective genome for each species using STAR aligner (v2.7.10) (37), followed by execution of RSEM (38) to perform the gene expression quantification step. Differential expression analysis was performed using the programming language R (v4.2.0) and the R package edgeR (v3.42.2) (39). Filtering of low-abundance genes was achieved by employing the filterByExpr function (edgeR) with default arguments. Finally, the weighted trimmed mean of *M*-values (TMM) normalization method and quasi-likelihood *F*-testing were utilized to assess abundance changes between conditions (absolute logFC > 0.58, FDR < 0.05).

TDMD-sRNA-Seq. Small RNA-Seq libraries from ZSWIM8 knockout and control samples were quality checked and pre-processed using FastQC, minion and cutadapt. miRNA quantification and differential expression analysis was conducted executing DIANA-mAP at default settings. Publicly available, analyzed, RNA-Seq datasets were retrieved and gene expression measurements of control samples (Transcripts-Per-Million, TPM > 1) were retained for each utilized cell-line. In case of replicates, TPMs were

averaged. Candidate TDMD miRNA–gene interactions were formed by integrating (i) significantly up-regulated miRNAs upon ZSWIM8 knockout (i.e. $\log_{2}FC > 0.585$ $FDR < 0.05$) and (ii) protein-coding genes with expression levels above the third quartile of expression in the same condition, which (iii) exhibited direct miRNA-MRE interactions, as evidenced by our in-house analyzed CLIP-Seq datasets.

Text-mining pipeline

A Natural Language Processing (NLP) pipeline was employed to extract experimentally supported miRNA–gene interactions from PubMed literature. Initially, 4571 sentences were manually labeled as positive (2722 sentences) or irrelevant (1849 sentences). Machine learning was then utilized, specifically employing a BioBERT (40) model for sentiment analysis to classify these sentences. Subsequently, the Stanza model (41) was used to extract dependency trees, and the Term Frequency - Inverse Document Frequency (TF-IDF) statistical method and cosine similarity were applied to capture the relationships between words, particularly those linking keywords and each other. This process facilitated the identification of essential terms in the context of miRNA and gene references, ultimately contributing to the determination of a threshold. This threshold enabled the accurate extraction of interactions from positive sentences. Additionally, the BERN2 tool (42) for Named Entity Normalization (NEN) was integrated into the pipeline. The methodology successfully processed approximately 39 000 papers retrieved through an advanced PubMed query. To assess the quality of the predictions of the NLP model, we manually curated 400 randomly selected abstracts from papers already processed by the method and created a set of miRNA–gene pairs (643 interactions – Supplementary Table 1). Using this set as ground truth, we calculated the True Positive (TP) and False Positive (FP) percentages of the text-mining-derived miRNA–gene interactions. The NLP model exhibited an ~67% TP rate (i.e. a 33% FP rate). To further refine the abstract-extracted interactions, we utilized a Large Language Model (LLM), specifically gpt-3.5-turbo-0301 and used it as an additional filter, applied to the NLP results. A refinement of the initial LLM results involved instructing the LLM to express the miRNA–gene relationships in its own words before translating them into a standardized format for enhanced comprehension. This additional step (LLM filter) yielded an 87.05% TP rate (i.e. a 12.95% FP rate). The total number of predicted interactions corresponds to approximately 19 000 entries originating from ~15 000 article abstracts.

Database architecture and implementation

TarBase-v9.0 is a single-page application (SPA) based on React.js (<https://github.com/facebook/react>) and Tailwind CSS (<https://github.com/tailwindlabs/tailwindcss>) frameworks. React Router (<https://github.com/remix-run/react-router>) handles the SPA's navigation and Recharts (<https://github.com/recharts/recharts>), Nivo (<https://github.com/plouc/nivo>) web frameworks are utilized for data visualization. The SPA interacts asynchronously with a Representational State Transfer Application Programming Interface (REST API), which was implemented in the Rust programming language (<https://www.rust-lang.org>), extending Axum (<https://github.com/tokio-rs/axum>) framework. The RESTful server communicates with a PostgreSQL

relational database (<https://github.com/postgres/postgres>) via the Diesel Object-Relational Mapper (ORM) (<https://github.com/diesel-rs/diesel>). The database contains TarBase data in Second and Third Normal Forms (2NF, 3NF) for a manageable number of tables. The queries are optimized by efficient table joining and indices on columns that participate in conditions. Also, the global statistics of the database are cached using Materialized Views.

Discussion

Meticulous production, curation and indexing of miRNA–gene interactions across various experimental contexts and species is essential for both fundamental and practical RNA research endeavors. TarBase-v9.0 highlights the ongoing commitment to systematically cataloging millions of experimentally supported miRNA targets for over 15 years. The latest version hosts ~6 million entries, comprising ~2 million unique miRNA–target pairs, supported by 37 experimental (high- and low-yield) protocols in 172 tissues and cell-types. Importantly, for the first time, it also lists interactions leading to target-directed miRNA degradation (TDMD) events and millions of miRNA-binding locations with cell-type resolution produced by the *de novo* analysis of state-of-the-art NGS methodologies. This content is provided to users through a new interface that introduces numerous new features and enhances its adaptability and user-friendliness.

TarBase v9.0 is the largest collection of experimentally supported miRNA–gene interactions to date. This wealth of information can enforce or even at cases substitute *in silico* predicted interactions.

TarBase may also be accessed here: <http://62.217.122.56/>.

Data availability

TarBase is freely available at <https://dianalab.e-ce.uth.gr/tarbasev9> and may also be accessed at <http://62.217.122.56/>. No new data were generated or analysed in support of this research.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Dr Theodore Dalamagas and Dr Thanasis Vergoulis at the 'Athena' Research Center in Athens, Greece for their assistance with the deployment of TarBase-v9.0.

Funding

European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE [T2EΔK-00391]; PhD scholarship, in the operating framework of the Center of Research Innovation and Excellence of University of Thessaly, funded by the Special Account for Research Grants of the University of Thessaly, is acknowledged. We also acknowledge the support of this work by the project 'Par-ICT CENG: Enhancing ICT research infrastructure in

Central Greece to enable processing of Big data from sensor stream, multimedia content and complex mathematical modeling and simulations' [MIS 5047244], which is implemented under the Action 'Reinforcement of the Research and Innovation Infrastructure', funded by the Operational Programme 'Competitiveness, Entrepreneurship and Innovation' [NSRF 2014-2020]; Greece and the European Union (European Regional Development Fund); Hellenic Foundation for Research and Innovation (H.F.R.I.) under the '1st Call for H.F.R.I. Research Projects to support Faculty Members & Researchers and the Procurement of High- and the procurement of high-cost research equipment grant' [2563]. Funding for open access charge: SCHINOS: clinical aSsessment of CHIos mastic effects in giNgivitis and Oral microbiome using advanced Sequencing techniques [T2EΔK-00391].

Conflict of interest statement

None declared.

References

- Liu, N. and Olson, E.N. (2010) MicroRNA regulatory networks in cardiovascular development. *Dev. Cell*, **18**, 510–525.
- Vo, N.K., Cambonne, X.A. and Goodman, R.H. (2010) MicroRNA pathways in neural development and plasticity. *Curr. Opin. Neurobiol.*, **20**, 457–465.
- Takaya, T., Ono, K., Kawamura, T., Takanabe, R., Kaichi, S., Morimoto, T., Wada, H., Kita, T., Shimatsu, A. and Hasegawa, K. (2009) MicroRNA-1 and MicroRNA-133 in spontaneous myocardial differentiation of mouse embryonic stem cells. *Circ. J.*, **73**, 1492–1497.
- Rottiers, V. and Näär, A.M. (2012) MicroRNAs in metabolism and metabolic disorders. *Nat. Rev. Mol. Cell Biol.*, **13**, 239–250.
- Peng, Y. and Croce, C.M. (2016) The role of MicroRNAs in human cancer. *Signal Transduct. Targeted Ther.*, **1**, 15004.
- Vasudevan, S., Tong, Y. and Steitz, J.A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science*, **318**, 1931–1934.
- Hwang, H.-W., Wentzel, E.A. and Mendell, J.T. (2007) A hexanucleotide element directs microRNA nuclear import. *Science*, **315**, 97–100.
- Tang, R., Li, L., Zhu, D., Hou, D., Cao, T., Gu, H., Zhang, J., Chen, J., Zhang, C.-Y. and Zen, K. (2012) Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. *Cell Res.*, **22**, 504–515.
- Turchinovich, A., Weiz, L., Langheinz, A. and Burwinkel, B. (2011) Characterization of extracellular circulating microRNA. *Nucleic Acids Res.*, **39**, 7223–7233.
- Shin, S., Park, Y.H., Jung, S.-H., Jang, S.-H., Kim, M.Y., Lee, J.Y. and Chung, Y.J. (2021) Urinary exosome microRNA signatures as a noninvasive prognostic biomarker for prostate cancer. *NPJ Genomic Med.*, **6**, 45.
- Tastsoglou, S., Miliotis, M., Kavakiotis, I., Alexiou, A., Gkotsi, E.C., Lambropoulou, A., Lygnos, V., Kotsira, V., Maroulis, V. and Zisis, D. (2021) Plasmir: a manual collection of circulating microRNAs of prognostic and diagnostic value. *Cancers*, **13**, 3680.
- Chen, X., Liang, H., Zhang, J., Zen, K. and Zhang, C.-Y. (2013) microRNAs are ligands of toll-like receptors. *RNA*, **19**, 737–739.
- Raden, M., Wallach, T., Miladi, M., Zhai, Y., Krüger, C., Mossmann, Z.J., Dembny, P., Backofen, R. and Lehnardt, S. (2021) Structure-aware machine learning identifies microRNAs operating as toll-like receptor 7/8 ligands. *RNA Biol.*, **18**, 268–277.
- Chi, S.W., Zang, J.B., Mele, A. and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature*, **460**, 479–486.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C. and Munschauer, M. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
- Moore, M.J., Scheel, T.K., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M. and Darnell, R.B. (2015) miRNA–target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.*, **6**, 8864.
- Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654–665.
- Gay, L.A., Sethuraman, S., Thomas, M., Turner, P.C. and Renne, R. (2018) Modified cross-linking, ligation, and sequencing of hybrids (qCLASH) identifies Kaposi's sarcoma-associated herpesvirus microRNA targets in endothelial cells. *J. Virol.*, **92**, e02138–17.
- Han, J., La Vigne, C.A., Jones, B.T., Zhang, H., Gillett, F. and Mendell, J.T. (2020) A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science*, **370**, eabc9546.
- Shi, C.Y., Kingston, E.R., Kleaveland, B., Lin, D.H., Stubna, M.W. and Bartel, D.P. (2020) The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science*, **370**, eabc9359.
- Huang, H.-Y., Lin, Y.-C.-D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J. and Zuo, H. (2022) miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res.*, **50**, D222–D230.
- Li, J.-H., Liu, S., Zhou, H., Qu, L.-H. and Yang, J.-H. (2014) starBase v2.0: decoding miRNA–ceRNA, miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
- Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X. and Li, T. (2009) miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
- Paraskevopoulou, M.D., Karagkouni, D., Vlachos, I.S., Tastsoglou, S. and Hatzigeorgiou, A.G. (2018) microCLIP super learning framework uncovers functional transcriptome-wide miRNA interactions. *Nat. Commun.*, **9**, 3601.
- Tastsoglou, S., Alexiou, A., Karagkouni, D., Skoufos, G., Zacharopoulou, E. and Hatzigeorgiou, A.G. (2023) DIANA-microT 2023: including predicted targets of virally encoded miRNAs. *Nucleic Acids Res.*, **51**, W148–W153.
- Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R. and Bhair, J. (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.
- Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
- Karagkouni, D., Paraskevopoulou, M.D., Chatzopoulos, S., Vlachos, I.S., Tastsoglou, S., Kanellos, I., Papadimitriou, D., Kavakiotis, I., Maniou, S. and Skoufos, G. (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.*, **46**, D239–D245.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.*, **17**, 10–12.
- Wu, T.D., Reeder, J., Lawrence, M., Becker, G. and Brauer, M.J. (2016) GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Methods Mol. Biol.*, **1418**, 283–334.
- Aparicio-Puerta, E., Hirsch, P., Schmartz, G.P., Fehlmann, T., Keller, V., Engel, A., Kern, F., Hackenberg, M. and Keller, A. (2023) isomiRdb: microRNA expression at isoform resolution. *Nucleic Acids Res.*, **51**, D179–D185.
- Kavakiotis, I., Alexiou, A., Tastsoglou, S., Vlachos, I.S. and Hatzigeorgiou, A.G. (2022) DIANA-miTED: a microRNA tissue expression database. *Nucleic Acids Res.*, **50**, D1055–D1061.
- Alexiou, A., Zisis, D., Kavakiotis, I., Miliotis, M., Koussounadis, A., Karagkouni, D. and Hatzigeorgiou, A.G. (2020) DIANA-mAP:

- analyzing miRNA from raw NGS data to quantification. *Genes*, **12**, 46.
34. Travis,A.J., Moody,J., Helwak,A., Tollervey,D. and Kudla,G. (2014) Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods*, **65**, 263–273.
 35. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA package 2.0. *Algorith. Mol. Biol.*, **6**, 1–14.
 36. Davis,M.P., van Dongen,S., Abreu-Goodger,C., Bartonicek,N. and Enright,A.J. (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods*, **63**, 41–49.
 37. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
 38. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.*, **12**, 323.
 39. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
 40. Lee,J., Yoon,W., Kim,S., Kim,D., Kim,S., So,C.H. and Kang,J. (2020) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, **36**, 1234–1240.
 41. Qi,P., Zhang,Y., Zhang,Y., Bolton,J. and Manning,C.D. (2020) Stanza: a Python natural language processing toolkit for many human languages. arXiv doi: <https://arxiv.org/abs/2003.07082>, 23 April 2020, preprint: not peer reviewed.
 42. Sung,M., Jeong,M., Choi,Y., Kim,D., Lee,J. and Kang,J. (2022) BERN2: an advanced neural biomedical named entity recognition and normalization tool. *Bioinformatics*, **38**, 4837–4839.